

Second Workshop on Subword and Character Level Models in NLP 2018

New Orleans, Louisiana, USA
6 June 2018

ISBN: 978-1-5108-6361-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2018) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2018)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Morphological Word Embeddings for Arabic Neural Machine Translation in Low-Resource Settings</i> Pamela Shapiro and Kevin Duh	1
<i>Entropy-Based Subword Mining with an Application to Word Embeddings</i> Ahmed El-Kishky, Frank Xu, Aston Zhang, Stephen Macke and Jiawei Han	12
<i>A Comparison of Character Neural Language Model and Bootstrapping for Language Identification in Multilingual Noisy Texts</i> Wafia Adouane, Simon Dobnik, Jean-Philippe Bernardy and Nasredine Semmar	22
<i>Addressing Low-Resource Scenarios with Character-aware Embeddings</i> Sean Papay, Sebastian Padó and Ngoc Thang Vu	32
<i>Subword-level Composition Functions for Learning Word Embeddings</i> Bofang Li, Aleksandr Drozd, Tao Liu and Xiaoyong Du	38
<i>Discovering Phonesthemes with Sparse Regularization</i> Nelson F. Liu, Gina-Anne Levow and Noah A. Smith	49
<i>Meaningless yet meaningful: Morphology grounded subword-level NMT</i> Tamali Banerjee and Pushpak Bhattacharyya	55
<i>Fast Query Expansion on an Accounting Corpus using Sub-Word Embeddings</i> Hrishikesh Ganu and Viswa Datha P.	61
<i>Incorporating Subword Information into Matrix Factorization Word Embeddings</i> Alexandre Salle and Aline Villavicencio	66
<i>A Multi-Context Character Prediction Model for a Brain-Computer Interface</i> Shiran Dudy, Shaobin Xu, Steven Bedrick and David Smith	72